Error in the Sciences:
Diagnosis, Prognosis, and Rectifying Measures

A workshop
Organizing committee: Marcel Boumans, Giora Hon, and Arthur Petersen
24 October – 28 October 2011
Lorentz Center, Leiden, the Netherlands

The aim of the workshop is to explore various practices of dealing with error to attain reliability, and to gain a deeper understanding of what error in science and its treatment entails. While the daily practice of empirical research, in and outside the laboratory, is dominated by dealing with all kinds of errors to increase the reliability of the results, there exists no general cross-disciplinary framework for dealing with errors. Various sophisticated procedures for the systematic handling of observational and measurement errors, and procedures for data-analysis were and still are being developed, but they all are fragmented and mainly developed to address specific epistemological and methodological problems within a particular scientific domain. The reason that a more general account is – still – lacking is that the kind of error to be corrected differs from case to case and depends upon the effects of many different conditions peculiar to the subject under investigation, the research design, and the equipment and/or models used, so is context dependent and field specific. The various practices of dealing with errors have developed their own separate methods and techniques, with only little cross fertilization. While these different methods are not likely to be integrated, solutions to their common problem – how to take account of reliability – may well be. That is, while contextual knowledge is not easily transmittable to different scientific domains, methods for achieving reliability may well have an over-arching feature.

Our aim is to develop such a general framework while doing justice to the idiosyncrasy of the circumstances in which errors arise. This means that beside existing statistical analyses of data which in measurement science is called Type A evaluation, we wish to discuss Type B evaluations of uncertainty. Examples of Type A evaluations are calculating the standard deviation of the mean of a series of independent observations; using the method of least squares to fit a curve to data in order to estimate the parameters of the curve and their standard deviations; and carrying out an analysis of variance in order to identify and quantify random effects in certain kinds of measurements. The underlying assumptions for legitimizing Type A evaluations are the availability of a large number of independent observations, equally trustworthy so far as skill and care are concerned, and obtained with instruments with known precision, which may apply to many experimental practices, but are too abstract for many empirical research practices outside the laboratory. Type B evaluations are based on scientific judgment using all of the relevant information available, which may include experience with, or general knowledge of, the behavior and property of relevant materials and instruments, manufacturer's specifications, models, data provided in calibration and other reports, and uncertainties assigned to reference data taken from handbooks. The importance of Type B evaluations is acknowledged in measurement science:

> The evaluation of uncertainty is neither a routine task nor a purely mathematical one; it depends on detailed knowledge of the nature of the measurand and of the measurement. The quality and utility of the uncertainty quoted for the result of a measurement therefore ultimately depend on the understanding, critical analysis, and integrity of those who contribute to the assignment of its value. (*Guide to the Expression of Uncertainty in Measurement*, ISO, 1993, p. 8)

The workshop will use case studies of research practices across a wide variety of scientific and practical activities, and across a range of disciplines (including experimental physics, econometrics, environmental science, climate science, industrial engineering, measurement science, and statistics), with the aim of integrating epistemologies and methodologies of treatments of error in scientific discourse. Such integration may be achieved through cross-disciplinary transfer of diagnosis, prognosis, and rectifying measures. To achieve this goal we will invite practitioners working in various scientific domains and disciplines as well as philosophers and historians of science, to discuss strategies of dealing with error in theory and in practice.

The challenge of this workshop lies in its truly interdisciplinary nature. On the one hand we need to explore practices of dealing with error in a specific scientific domain; therefore we need practitioners to provide us with such details. On the other hand, we need to generalize from these detailed case studies to make comparisons between the different practices; therefore we need the philosophers and historians.

The workshop's *goal* is to develop an account of Type B evaluations. Although an account of Type A evaluation does exist, often denoted by 'theory of error', an account of Type B evaluation is still lacking. The reason for this is the only quite recent scientific interest in and acknowledgement of this (more subjective) aspect of errors. The organizers of this workshop have the experience to bring this workshop to a successful result, that is, a first outline of such account, published as a collection of high quality papers to fill a special issue of a journal or an edited volume.

Organizing committee

Marcel Boumans (University of Amsterdam) was organizer of a workshop 'Measurement in Economics', held in Amsterdam (April 2006). The aim of this workshop was similar to the aim of the proposed workshop, namely, to develop a framework to account for fragmented strategies of empirical research in various separate scientific domains with their own methodology and history. Participants were philosophers, historians, economists, econometricians, and a metrologist who discussed measurement practices in economics. The papers were published in an edited volume, *Measurement in Economics: A Handbook* (Elsevier, 2007). One of Boumans's main line of research since the early 1990s is the assessment of reliability of empirical research outside the laboratory.

Giora Hon (University of Haifa) was co-organiser (with Jutta Schickore and Friedrich Steinle) of the conference 'Going Wrong and Making it Right: Error as a Crucial Feature of Concept Adjustments in Experimental Contexts', held in Aegina, Greece, Spring 2003. The papers presented by philosophers and historians of science at the conference were recently published in a volume edited by the organisers, *Going Amiss in Experimental*

Research (Springer, 2009). The proposed workshop can be considered an extension of this conference in the sense that it now also will include non-experimental research to which also practitioners will contribute. Hon was also co-organiser (with Vincent Icke and James McAllister) of the Lorentz workshop 'Symmetry as a Modern Scientific Concept: Historical and Philosophical Perspectives' (11–14 March 2008). Since his dissertation (University of London, 1985) 'On the Concept of Experimental Error', he has published widely on the problem of error from philosophical and historical perspectives (see especially his paper, 'Towards a Typology of Experimental Errors: an Epistemological View.' *Studies in History and Philosophy of Science* 20 (1989) 469–504).

Arthur Petersen is Director of the Methodology and Modelling Programme at the Netherlands Environmental Assessment Agency, NL. He received graduate training as theoretical physicist (MSc), atmospheric scientist (PhD) and philosopher of science (MA and PhD). Since 2001, he has gained considerable experience in shaping the science-policy interface at Dutch, European and global levels on issues of climate change and sustainable development. He has become a world-leading expert on assessing and communicating uncertainties. In particular, he has studied major uncertainties in climate simulation. Relevant publication: *Simulating Nature: A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice* (dissertation), 2006.

Error in the Sciences: Diagnosis, Prognosis, and Rectifying Measures

DAY 1 Monday:

[I] Historical and Philosophical Perspectives on Error in Science

Traditionally history and philosophy of science has been concerned with practices that claim to attain (scientific) knowledge. Although always presents, error as an object of inquiry has been neglected, considered unproductive. This attitude is now changing. There is a growing interest in the concept of error, in its ramifications, and in the wider meaning of uncertainty. Day 1 will have two distinct parts: an introductory session to open the workshop and a focused discussion on handling error in experimentation. The introductory talk will map the problem of error in a historical and philosophical context. Different perspectives and topics will be presented in an overview: error in experimentation, error as historiographical problem, error statistics, and the like. A plenary session will follow in which contributors will present themselves with very short prepared statements of their contributions and the goals of the workshop will be discussed. The workshop will then continue with a session on error in experimentation. Like any goal-oriented procedure, experiment is subject to many kinds of error. They have a variety of features, depending on the particulars of their sources. The identification of error, its source, its context, and its treatment shed light on practices and epistemic claims. Understanding an error amounts, *inter alia*, to uncovering the knowledge generating features of the system involved—the very features that are the object of study of the historian-philosopher when it comes to evolving systems in scientific practice.

09.00 – 10.00	Arrival and registration
10.00 – 10.15	Introduction by the Lorentz Center staff and the organizing committee: **Marcel Boumans**, **Giora Hon**, and **Arthur Petersen**
10.15 – 11.15	**Giora Hon** and **Jutta Schickore**: *Introduction to the workshop*
11.15 – 11.45	Coffee break
11.45 – 12.30	Plenary session, **Giora Hon** (Chair of the Day): Formulating questions and aims for the workshop.
12.30 – 14.00	Lunch break

[II] Error and the Method of Experimentation (30 min. talk, 15 min. discussion)

14.00 – 15.30	*Observational Errors* *Challenges to Robustness*
15.30 – 16.00	Tea break
16.00 – 17.30	*Error Statistics and Experimentation*

Confounding Variables

17.30 – 19.00 Wine and cheese welcoming party

DAY 2 Tuesday: Measurement Errors

Measurement results are generally not considered as reports directly about the state of the object under measurement, but on our knowledge about this state. Measurement shifted from a truth-seeking process to a model-based one in which the quality of the measurement is assessed by pragmatic aims. As a result of the epistemological shift, the quality of measurement is not reported in terms of accuracy, an expression of closeness to the true value, but in terms of uncertainty. This has also had implications on calibration strategies: instead of expecting that reference values are true they are required only to be traceable. On Day 2 these shifts will be discussed by focusing on key issues: the shift from error to uncertainty, the shift from accuracy assessment to quality assessment and the shift from standards as prototypes to standards as instrumental set-ups.

09.15 – 09.30	**Marcel Boumans** (Chair of the Day): *Introduction to today's topic*
09.30 – 10.30	*Uncertainty Instead of Error*
10.30 – 11.00	Coffee break
11.00 – 12.00	*Quality Assessment*
12.00 – 13.30	Lunch break
13.30 – 14.30	*Standards*
14.30 – 15.30	**European Fusion Development Agreement** (EFDA)
	Report of the working group on data analysis: diagnostics
15.30 – 16.00	Tee break
16.00 – 17.00	Plenary discussion led by the Chair of the Day

DAY 3, Wednesday: Communicating Uncertainties

In science the need to communicate with decision makers about the uncertainties in the relevant models has become acute. Interdisciplinary work has been done in this domain to arrive at commonly agreed upon typologies of uncertainty. This includes efforts to widen the concept of reliability, since it is often not possible to establish the accuracy of the results of simulations or to quantitatively assess the impacts of different sources of uncertainty. On Day 3 recourse will be made to qualitative assessment of the different elements used in the research (e.g., data, models, expert judgments and the like) and determine their "methodological reliability", given the purpose of the relevant model.

09.15 – 09.30	**Arthur Petersen** (Chair of the Day): *Introduction to today's topic*

09.30 – 10.30	*Chaos and Model Uncertainty in Forecasts and Projections*
10.30 – 11.00	Coffee break
11.00 – 12.00	*Foresight and Models*
12.00 – 13.30	Lunch break
13.30 – 14.30	*Modelling and Evidence*
14.30 – 15.30	*Risk Analysis Methodology, Uncertainty Analysis, and Expert Judgment*
15.30 – 16.00	Tea break
15.30 – 17.00	Plenary discussions led by the Chair of the Day

PUBLIC SESSION (in town)

Communicating scientific results to lay audience, facts or uncertainties (the case of climate change)

WORKSHOP DINNER (in town)

DAY 4 Thursday: Social Science and Statistics

Data of social science and statistics are typically inhomogeneous: as the realizations of complex interactions they are not stable. Since the traditional statistical techniques presuppose homogeneity, they cannot be applied in these instances. Various 'ometrics'-disciplines arose as new branches of applied statistics by developing strategies to treat this kind of data. The new strategies share the feature of being model based. An evaluation of errors therefore is a model-based assessment, where the model must cover the sources of errors. On Day 4 several strategies will be discussed where errors are evaluated by the assessment of its representations.

09.15 – 09.30	**Marcel Boumans** (Chair of the Day): *Introduction to today's topic*
09.30 – 10.30	*Errors and Bias in Index Numbers*
10.30 – 11.00	Tea break
11.00 – 12.00	*The Role of Errors in Statistical Models*
12.00 – 14.30	Lunch break
14.30 – 15.30	*Sensitivity*
15.30 – 16.00	Tea break
16.00 – 17.00	*Uncertainties in the Social Sciences* (panel discussions, led by the Chair of the Day)

Day 5 Friday: Concluding session: Lies, damned lies, and statistics

The theme of use and misuse of statistics in science and the public domain concludes the workshop. Participants summarize and evaluate the results of the workshop; plans for future collaborations and new researches are discussed.

09.15 – 09.30	**Giora Hon** (Chair of the Day): *Introduction to today's topic*
09.30 – 10.30	*The Use of Statistics in Public Domains*
10.30 – 11.00	Coffee break
11.00 – 12.00	*Error in Social Science*
12.00 – 13.30	Lunch break
13.30 – 15.00	Concluding sessions, plans for the future

END